DOCUMENT RESUME

ED 091 408 TM 003 619

AUTHOR Angrist, Shirley S.; Borke, Helene
TITLE An Observational System for the Evaluation of Preschool Programs: Research and Development.
SPONS AGENCY Alcoa Foundation, Pittsburgh, Pa.
PUB DATE [Apr 74]
NOTE 20p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, Illinois, April, 1974)

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS *Classroom Environment; *Classroom Observation Techniques; Cognitive Processes; Day Care Services; Evaluation Techniques; *Formative Evaluation; Instructional Improvement; Interpersonal Competence; *Preschool Programs; Program Evaluation; Student Behavior; Student Teacher Relationship; Teacher Behavior

ABSTRACT

An observational monitoring system is described for evaluating preschool programs and providing feedback to relatively untrained caregivers in order to encourage program improvement in the social and cognitive realms. The system utilizes a simple one page observation form for teachers and children detailing interaction and type of activity. Data are collected through accumulation of instantaneous observations with high interobserver reliability. Results show strong differences among three distinctive urban preschool settings, especially in teachers' behavior and children's sociometric patterns. The system's stress on simplicity and feedback should serve as a widely applicable formative evaluation tool for day care and nursery school settings. (Author)

ED 091408

TM 003 619

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

# AN OBSERVATIONAL SYSTEM FOR THE EVALUATION OF PRESCHOOL PROGRAMS: RESEARCH AND DEVELOPMENT

Shirley S. Angrist

School of Urban and Public Affairs

Carnegie-Mellon University

Helene Borke

Hay Associates

Paper presented at the American Educational Research Association annual meeting, Chicago, April 1974.

This research was supported by a grant from the Alcoa Foundation.

ABSTRACT

An Observational System for the Evaluation of Preschool Programs: Research and Development

Shirley S. Angrist, Carnegie-Mellon University

Helene Borke, Hay Associates

This paper describes the pilot study of an ongoing effort to develop an observational monitoring system which can tap program quality in preschool settings, including nursery schools and day care centers. The research objective is to devise simple techniques usable by teachers or caregivers for assessing the level of social and cognitive content and for providing feedback to teachers so that they can alter and improve programs. The monitoring system consists of one page forms for observing children and teachers in terms of interaction, intent of the interaction, and the extent of cognitive, language, creative or motor activity. In addition, the system includes a manual for training observers and a resource book of program materials for teachers. The observations were collected over six weeks through an accumulation of instantaneous photographic-like pictures for each child and teacher. This method yields high reliability with inter-observer agreement rates of over 90 percent. The data based on 20 observations per child and 30 per teacher were analyzed using analysis of variance and canonical analysis to ascertain school differences. The observational system was tested in 3 urban preschool settings. The results show that the three schools, while similar in fostering constructive activity, differ both in teacher and child behaviors. Significant differences occur in the extent of teachers' interaction with children, prosocial behavior and creativity, as well as in the amount of children's involvement, interaction with other children, cognitive and fine motor activity. This monitoring system, while still under development, is important in answering questions about the situational and organized contexts which foster social and intellectual competence in children. The implications of this research are twofold: (1) providing effective methods for monitoring the environmental dimensions of early childhood educational programs, so that caregivers and parents can understand the process and implement desired changes in the ongoing system; and (2) training caregivers to create and sustain intellectually stimulating and socially constructive child care environments.

The growing interest in early childhood education outside the family reflects the simultaneous recognition that family functions are changing in the U.S. and that developmental environments are necessary for young children. In addition to programs which emphasize play and social development there are a number of programs specifically designed for educating preschool children. These include the Montessori method, Piagetian-based learning, the open classroom, and Bereiter and Engleman's skills development approach. One of the major questions still to be answered is what effect these various programs have on the children attending them.

How can we measure what happens in preschool settings? Can we gauge the educational quality of the environment and discriminate good programs from bad ones? The purpose of the research reported in this paper is to develop an observational monitoring system which can tap program quality (regardless of orientation) in preschool settings. Our aim is to devise simple techniques usable by outside observers, teachers or caregivers, for assessing the level of social and cognitive functioning and for providing feedback to teachers so that they can alter and improve programs.

One of the main strategies emphasized by educational researchers has been to assess knowledge of subject matter, acquired as a result of a particular educational intervention using measures such as reading readiness or school achievement. Such approaches have been critized recently because they use summative measures rather than formative ones (Bloom, 1971) and because they frequently focus on inappropriate dimensions of program quality (Shapiro, 1973). Measures which dwell on school achievement ignore social and interpersonal competencies, as well as creative and problem-solving skills. In addition, achievement scores available "after-the-fact", are unusuable for changing an ongoing program. Bloom (1971) advocates a formative approach

using a variety of measures to evaluate actual classroom activities. Especially in preschool settings, formative evaluation may be the method of choice in order to provide feedback to teachers (Kamii, 1971; Shapiro, 1973).

The concern with evaluation in early childhood education is reflected in the proliferation of observation systems for the preschool and primary elementary school levels. The Analysis of Communication in Education (Bowman, 1972), The Differentiated Child Behavior Observation Classroom Scan (Ross and Zimiles, 1973), as well as the set of observational instruments developed at the Stanford Research Institute (Stallings et. al., 1973) reflect efforts to assess Follow Through programs, the sequel to Head Start. Other observational systems, like PROSE (Medley et. al., 1971) grew out of the concern to systematically describe the classroom context.

The varied observational instruments utilize a mixture of methodologies: some concentrate on classroom climate as reflected in teachers' behavior (Withall, 1949), others focus primarily on teachers in relation to children (Medley et. al., 1971 and Brandt, 1972), still others try to picture the whole room or situation (Ross and Zimiles, 1973). Observational techniques include instantaneous snapshot-like methods and longer time scans.

Our system builds on existing instruments but with the aim of creating a flexible system which is easy to use, yet one which highlights the key elements of "adequate" preschool programs. We began with Kamii's (1971) assumptions that a good program should contain social, emotional, cognitive, perceptual-motor and creative components. Hence our instruments assess social behavior in interactions, participation patterns, the use of language, mathematical and general knowledge, fine and gross motor coordination, and creative activities.

In addition we sought a system (1) flexible enough to assess situational or group context, on the one hand, and individual children or teachers' behavior, on the other hand; (2) capable of focusing predominantly on children or on teachers through parallel but separate forms; (3) yielding high reliability; (4) producing easily computerizable data; (5) and simple enough to be used by teachers for their own feedback.

## The Setting

The observational system was tested in three urban preschool settings. The three schools represent different spousorship, socioeconomic levels and program philosophy.

School I is a self supporting university based day care center. Fees are charged on a sliding scale. The program is oriented around play and social development. The 31 youngsters attending the school are predominantly children of university students but also of staff, faculty and community people. The group is mainly white with 25% from nonwhite backgrounds. The children are divided into 2 groups according to age; each group has about 2/3 boys and 1/3 girls. The ratio of children to teachers, is 15 to one. If aides are included the child-adult ratio is 8 to one.

School II is an OEO subsidized day care center. The program combines free play and the use of structured kits to teach cognitive skills. Only children of poverty families are eligible to attend under Title IV A. The 30 children are separated into two classrooms with 3 and 4 year olds in each room. All the youngsters are black and 60% are boys. As in School I, the child-teacher ratio is 15 to one. If aides are included the child to adult ratio is 8 to one.

School 3 is a university sponsored preschool and kindergarten. Fees are comparable to those of other private nursery schools. The program emphasizes an open classroom individualized approach and encourages cognitive growth. One fourth of the children are youngsters of faculty members with the remainder coming from the general community. The 3, 4 and 5 year olds are all combined into one large group of approximately 30 youngsters. Half the children are boys, half girls and 15% are from non-white backgrounds. The child-teacher ratio is 10 to 1 but drops to about 5 to 1 when students function as teaching aides.

## Procedures

The monitoring system involves four components: observations forms for teachers, forms for children, a manual for instructing observers, and resource materials for use by teachers. These components may be used separately or in combination. Copies of the observation forms and the manual are available from the authors.

Each form provides data on interaction, including who interacts with whom, the intent of the interaction, and type of activity in terms of language or other cognitive, creative or motor content. To collect the data, the observer first gets acclimated. Then acting like a photographer -- at the sound of a signal from the tape recorder, the observer notes what the observed child or teacher is doing at that instant and then completes the form. Observers randomly select which child or teacher to observe until all in the school have been observed at least once before making a second observation. Observations were made over two months in three schools until we had 30 observations per teacher and 20 per child.* All observations were made during the morning

---

*Two trained observers observing about 3 hours a day can complete about 2 observations for each of 30 children and about 4 observations on each of 3 teachers. We found that with absences, class trips and other special activities, the accumulation of 600 observations for children and 100 for teachers required at least 3 to 4 weeks in each school.

since in most programs this is the key indoor program time. Morning and indoor activity seemed to reveal best the programmatic emphasis of the school: activities were usually organized by the teachers, or free play or transitional. At other times of the day more time was spent in meals, naps and outdoor play. The data reported in this paper stem from observations of individual children and teachers.

Reliability was assessed in two ways: interobserver agreement and stability over time. For both children and teachers, observers worked in pairs. A realiability check was made on every tenth observation; observers synchronized these observations to respond to the same sound signal and the same subject in the same context. The interobserver agreement on observational categories for teachers' behavior was 97 percent, for children, 95 percent.

Stability of behavior over time was assessed through comparing two sets of teacher observations: 30 observations taken over a 4-6 week period were compared with 20 additional observations per teacher taken in one day. Ten out of 12 chi square tests of the distributions of teacher behaviors under each condition were not statistically significant. Hence we conclude that the teacher form yields rather stable patterns over time.

We reviewed the validity of the behavior categories in several ways. First, we modified or eliminated categories which, even after careful training, observers interpreted differently. Second, we tested to see whether other variables such as sex or age might strongly affect the behavioral categories. There were no differences at all between boys and girls in any of the schools on any of the categories. In addition, we checked for the possibility that School III children might score higher because some of them are five years old, while the other two schools have only three and

four year olds. Comparison of the five year olds with the three's and four's in School III yielded no significance differences in chi square tests of the incidence of each behavior category. Finally, we presented teachers in two schools with observational results based on earlier versions of the instruments and modified or eliminated behavioral categories according to their reactions about confusing, unclear or invalid data.

## Results

Our hypothesis that the monitoring instruments should detect differences in programmatic orientation among the three preschools was generally borne out, but with some qualifications.

First, we discovered some surprising similarities among the schools. Children were constructively involved in activities 90 percent of the time they were observed. Rarely (about 10 percent of the time) were children in any of the schools found to be bored, sad, angry or crying. Overall interaction rates among adults and children were fairly similar -- children were observed in interactions of some kind about half the time. Perhaps because the data were collected indoors, gross motor activity represented only about 15 percent of the morning's activities in each school.

Even for teachers, there were some school similarities: disciplining of children by ostracism, shouting or physical punishment was rare. But also rare were efforts to clarify social interaction or to express affection.

In order to test the hypothesis that the three schools should differ in teacher and child behavior, we used analysis of variance and canonical correlation techniques. For the analyses of variance we treated each observation as a discrete unit of data and converted each observation to

a score from 0 - 2, with 0, if the event did not occur, 1, if the event occurred once, and 2, if the event occurred twice per observation.

Table 1 presents the summary of the analysis of variance results for the three teachers in each school (with 30 observations per teacher).* On all the teacher variables, School III scores higher than School II, and School II scores higher than School I. School III teachers interact with children, express prosocial intent in their interactions with children and are involved in creative activity with children significantly more than the teachers in Schools I and II. School III teachers are also highest on cognitive and gross motor activity, but these differences do not achieve statistical significance.

When we turn to the child behavior variables presented in Table 2, it is evident that School I children tend to score higher than School II children in all categories except for fine motor activity. This confirms what our observers noted that School II was well supplied with puzzles compared with School I which had few toys requiring fine motor skills. In contrast with the other two schools, School III children score highest on all behaviors. They are significantly more involved overall, they interact with other children more, they exhibit more prosocial behavior, and they spend more time in cognitive activity.

To determine the extent to which the schools differ in overall programming, we used canonical analysis with the computer statistical package developed

---

*For all schools we omitted observations on aides from these analyses because their roles and numbers vary greatly, and because we assumed that teachers are the main determiners of the school program.

by Gleason (1973). Only the behavioral categories which reached statistical significance in the analyses of variance were included. Observations on each behavior category were summed separately for teachers and children for each morning of the data collection period. The bulk of the data was collected in 15 days at each school. It is these 15 sets of "morning observation scores" which served as the input for the canonical analysis.

The canonical analysis yields two variates. Table 3 presents the structure matrix for each variate. These are the correlations of each behavioral and school variable to each variate and are interpretable as factor loadings in factor analysis (Stewart and Love, 1968). The first variate is dramatically most characteristic of School III and least characteristic of School I. School III is high on children's participation in cognitive activity. It is also high on the amount of teacher interaction with children, on teachers' expression of prosocial intent and on teachers' encouragement of creativity. Variate B best describes School II as very high and School I as extremely low in children's fine motor activity. With 50 percent of the variance in schools explained by the behavioral categories, the results clearly suggest important school differences in program content. It is interesting to note that in School III both children's and teacher's behaviors serve to highlight program differences.

Figure 1 presents the canonical analysis results graphically, showing the extent to which the behavioral variables occupy the position in space covered by the schools. With Variate A on the horizontal axis and Variate B on the vertical axis, we can see that School III dominates the graph surrounded by the cluster of both child and teacher variables. Seven of the eight variables clearly characterize School III and tend to be diametrically

opposite descriptors of Schools I and II. The teacher variables are closest to School III; the first three child variables are also nearest to School III but they are somewhat closer than any other variables to School I. School II is in reasonable proximity to variable 5, children's fine motor activity. Schools I and II appear rather isolated spatially from School III. We can infer that in School III, the teachers' emphasis on interaction with children and their manifestation of prosocial behavior encourages the children to interact and to express prosocial intent. Figure 1 confirms the greater developmental quality of School III's program.

## Discussion

The results of the observational data presented so far indicate that the observation system does tap empirical differences in schools which are characterized as having different programs. While this fulfills our primary goal to develop simple observational instruments for assessing school quality, there are other important considerations which justify these instruments as part of a monitoring system.

Ideally, when it is fully operative, the system makes possible the following:

(1) feedback to teachers about their own behavior as individuals or as a group.

(2) feedback to teachers about children either as individuals or as a group.

(3) a picture of changes in teachers' or children's behavior over time, or due to program alterations.

Once the observers are trained and the observational data are collected, key punched and stored on a computer, the evaluators can feed back information about each behavior category using data derived from simple frequency distributions.

Since this research began, we have had the opportunity to determine the instruments' ability to detect program changes within a school. By chance, two of the three schools underwent significant alterations in staffing, size and equipment. Our preliminary results indicate that the instruments do reveal changes in the behavior categories consequent to these program alterations.

In a preliminary trial of the system, the investigators stored the data in an SPSS file (Nie, Bent and Hull, 1970) generated marginal distributions for each behavior category and then presented the results for individual children and for the whole school to teachers. Figure 2 has examples of the visually attractive and quickly interpretable results on two behavior categories as they were presented to teachers. Each circle represents 100 percent or all observations in that behavior category. The darkened area indicates the percent of occurrence of the given behavior. The teachers reacted by commenting about the individual children and by reassessing some children's behavior. We strongly recommend 40 observations per subject when providing teachers with feedback on an individual.

In order to reduce the time needed to collect observations in a school setting, we are extending the monitoring system by developing group or situationally oriented observation forms. With the situation observation forms, observers count the number of children in each activity category through a series of successive instantaneous observations. Observations in one classroom can be accumulated rapidly and reduced by computer to provide faster feedback to teachers. Since the situation forms, as well as the individual child and teacher forms center on the same activity dimensions, the system can serve in formative evaluation in two key ways:

(1) for contextual assessment of school program in terms of predominant cognitive and social activities involving children and teachers, and

(2) for behavioral assessment of individual teachers or children in order to train teachers in dealing with specific children or for altering the teacher's own behavior.

## Conclusion

Although our efforts to develop and refine the monitoring system are far from complete, we are encouraged by the results reported in this paper. The system's assets lie in its ability to detect the crucial cognitive and prosocial components of generally good but rather diverse preschool programs, its applicability for reflecting individual as well as overall school or classroom behavior, and its adaptability for feedback to teachers and program planners. The system's primary limitation lies in the time delay between the observational data collection to the feedback stage. A computerized information system could easily be developed to take the precoded observation sheets, score them by optical scanner and print out summary results.

TABLE 1. Summary of Analysee of Variance for Observations on Teacher Behaviors

| Variable | School Mean Score I (N=90) | II (N=90) | III (N=90) | F-Ratio |
|---|---|---|---|---|
| Interacts with children | .54 | .61 | .81 | 8.05** |
| Shows prosocial intent | .14 | .21 | .37 | 6.65** |
| In cognitive activity | .16 | .17 | .29 | 2.81 |
| In creative activity | .02 | .08 | .14 | 4.61* |
| In fine motor activity | .07 | .10 | .14 | 1.47 |

* Differences among schools are statistically significant at $p > .05$

** Differences among schools are significant at $p > .01$

For all analyses, df = 2/267.

TABLE 2. Summary of Analyses of Variance for Observations on Child Behaviors

| Variable | School Mean Score I (N=545) | II (N=600) | III (N=555) | F-ratio |
|---|---|---|---|---|
| Child is involved | .92 | .91 | .95 | 3.25* |
| Interacts with children | .29 | .21 | .29 | 5.92** |
| Interacts with adult | .23 | .20 | .25 | 1.97 |
| Shows prosocial intent | .33 | .27 | .38 | 8.42** |
| In cognitive activity | .19 | .13 | .28 | 22.06** |
| In creative activity | .18 | .16 | .18 | .92 |
| In fine motor activity | .14 | .28 | .23 | 16.43** |

* Differences among schools are statistically significant at $p > .05$

** Differences among schools are significant at $p > .01$

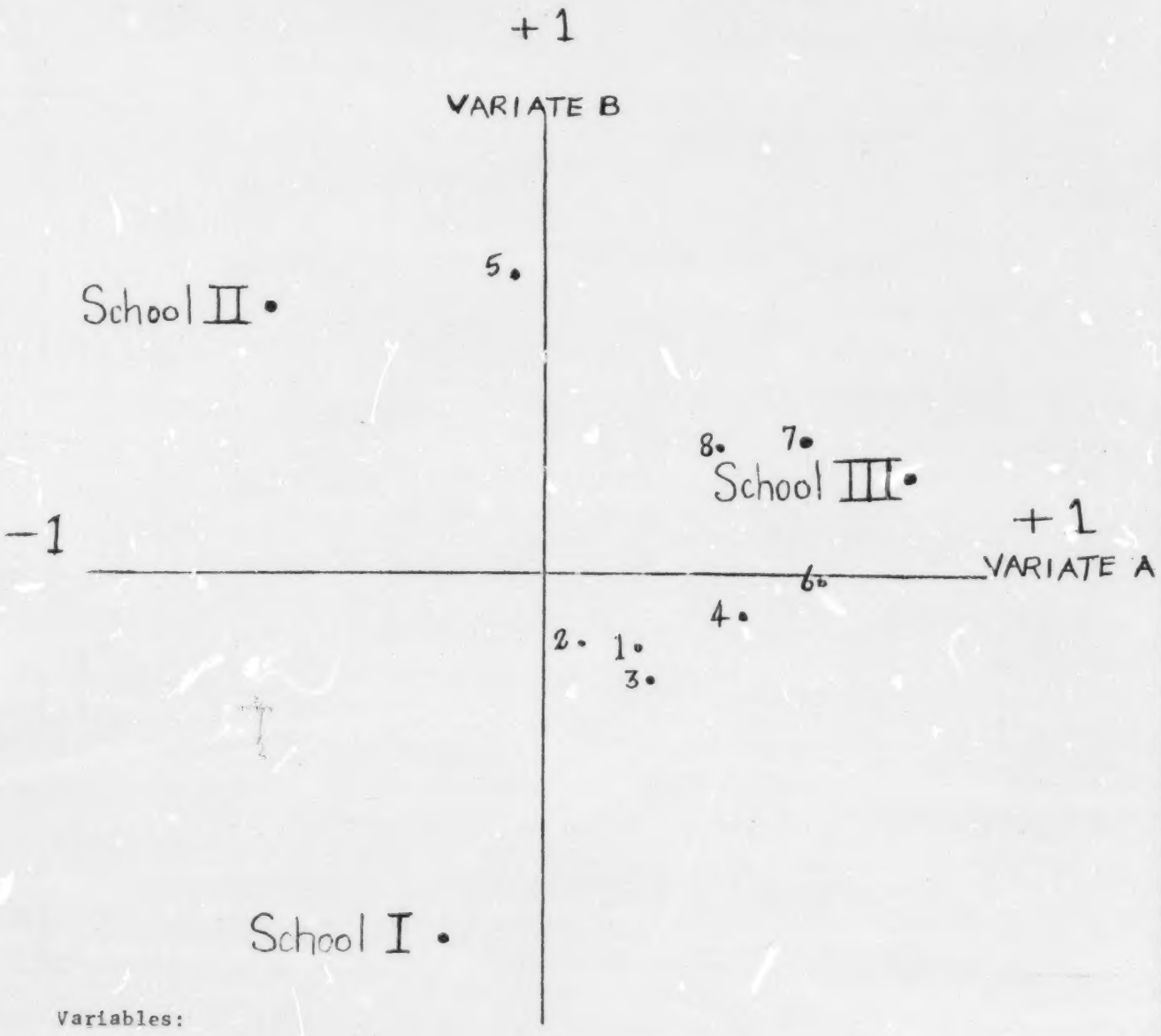
For all analyses, df = 2/1697.

TABLE 3. Results of Canonical Analysis

| | Structure Matrix of Canonical Coefficients | |
|---|---|---|
| | Variate A | Variate B |
| For Behavioral Categories: | | |
| 1. Child is involved | 0.25 | -0.19 |
| 2. Child interacts with children | 0.10 | -0.19 |
| 3. Child shows prosocial intent | 0.27 | -0.27 |
| 4. Child in cognitive activity | 0.54 | -0.12 |
| 5. Child in fine motor activity | -0.09 | 0.79 |
| 6. Teacher interacts with children | 0.73 | -0.01 |
| 7. Teacher shows prosocial intent | 0.69 | 0.35 |
| 8. Teacher encourages creativity | 0.46 | 0.34 |
| For Schools: | | |
| School I | -0.26 | -0.97 |
| School II | -0.71 | 0.71 |
| School III | 0.96 | 0.26 |
| | $R_{c1}$ = C.81 | $R_{c2}$ = 0.59 |

FIGURE 1. Program Characteristics of Three Schools:

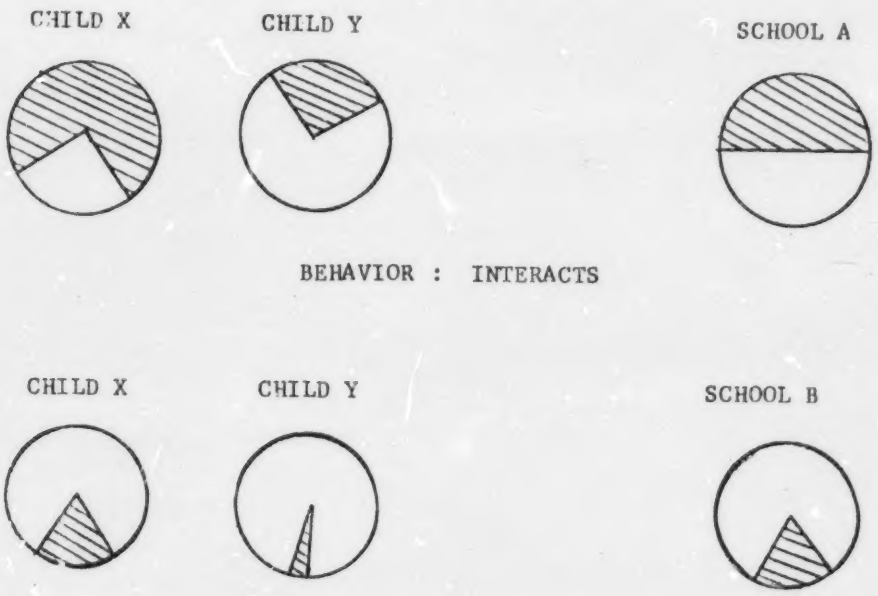
Structure Matrix for Two Canonical Variates




Variables:

1 = child is involved
2 = child interacts with children
3 = child shows prosocial intent
4 = child in cognitive activity
5 = child in fine motor activity
6 = teacher interacts with children
7 = teacher shows prosocial intent
8 = teacher encourages creativity

FIGURE 2. Graphic Presentation of Observations in Selected Behavior Categories for Individual Children and for Schools.



REFERENCES

Bloom, Benjamin S., Hastings, J. T., and Madaus, G. F. Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill, 1971.

Bowman, Garda W. Analysis of communication in education (ACE). New York: Bank Street College of Education, 1972.

Brandt, Richard M. An observational investigation of instruction and behavior in an informal British infant school. Charlottesville: University of Virginia, School of Education, 1972.

Gleason, Terry C. Some notes on canonical analysis. Pittsburgh: Carnegie-Mellon University, 1973.

Kamii, Constance. Preschool education: socioemotional, perceptual motor, cognitive development. In B. Bloom, J. T. Hastings and G. F. Madaus (eds.), Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill, 1971, 284-344.

Medley, Donald M., Quirk, T. J., Schluck, C. G. and Ames, N. P. The Personal Record of School Experiences: A Manual for PROSE Recorders. Princeton, New Jersey: Educational Testing Service, 1971.

Nie, Norman H., Bent, D. H., Hull, C. H. Statistical Package for the Social Sciences. New York: McGraw-Hill, 1970.

Ross, Sylvia and Zimiles, Herbert. The differentiated child behavior observational system. Paper presented at the American Educational Research Association annual meeting, New Orleans, 1973.

Shapiro, Edna. Educational evaluation: rethinking the criteria of competence. School Review, 1973, 81, 524-549.

Stallings, Jane, Steinmetz, Gerald and Wilcox, Mary. Observational findings for differences in Follow Through and Non-Follow Through classrooms. Paper presented at Society for Research in Child Development, Philadelphia, 1973.

Steward, D. and Love, W. A general canonical correlation index. Psychological Bulletin, 1968, 70, 160-163.

Withall, John. The development of a technique for the measurement of social emotional climate in classrooms. Journal of Experimental Education, 1949, 27, 347-361.